6.1: Probability Distributions Are Models of Random Experiments

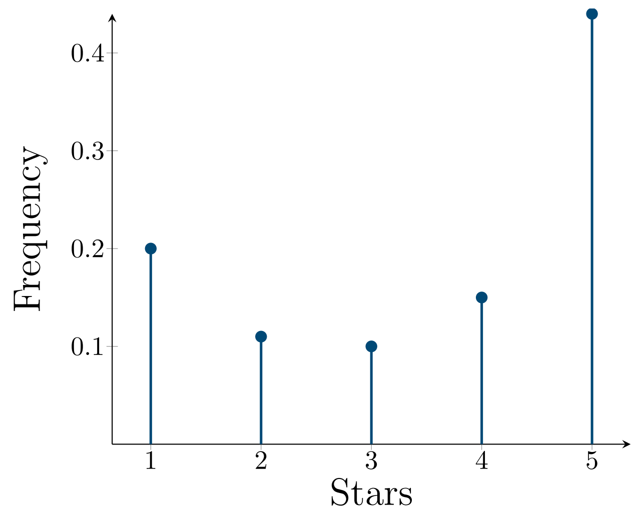
Definition.
The **probability distribution** describes

- all possible outcomes of a random experiment, and
- the probability of each outcome.

This is sometimes also referred to as the **probability distribution function (pdf)**.

Example. Suppose we are reading Amazon reviews of a particular product. In total, the product has $3,901$ reviews, distributed as shown below.

Stars	Frequency
5	0.44
4	0.15
3	0.10
2	0.11
1	0.20



If we pick a reviewer at random, what is the probability that they give a 5 star review? What about a 1 star review? $P(X=5)=0.44$

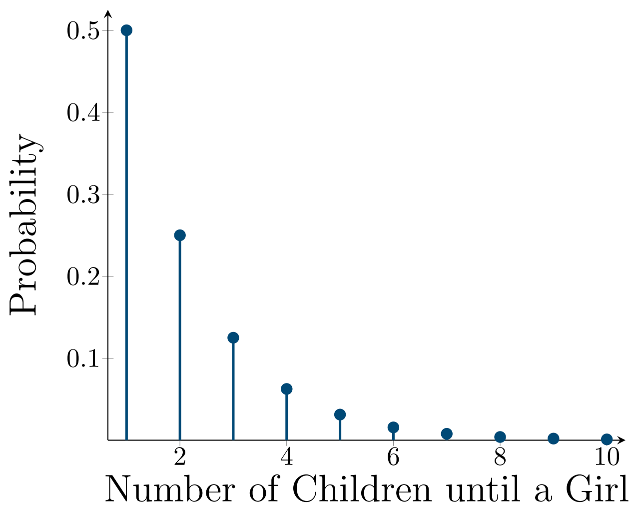
$P(X=1)=0.20$

What is the sum of the probabilities?

$0.44+0.15+0.10+0.11+0.20=1.00$

> *Note*: Valid probability distributions:
>
> - Have probabilities between 0 and 1,
> - The sum of the probabilities is *exactly* 1.

Example. Suppose a couple decides they will keep having children until they have a girl. Assuming that the likelihood of having a boy or girl is equally likely, the probability of having x children can be given by $(1/2)^x$, and is represented by the graph below.



What is the maximum number of children possible?

∞

Do the probabilities sum to 1?

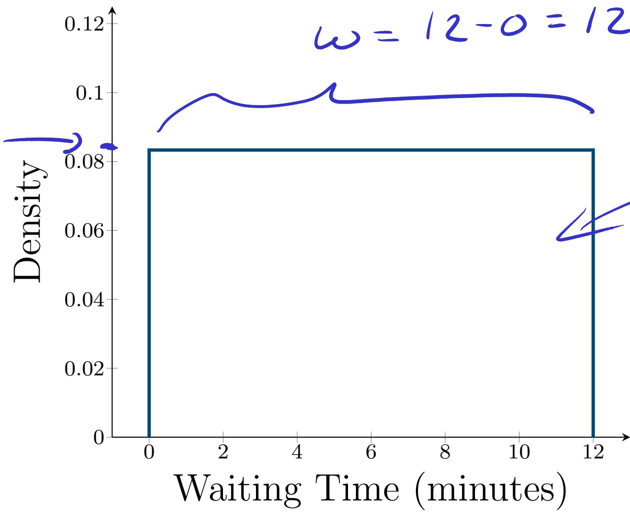
$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots + \frac{1}{1024} + \frac{1}{2048} + \dots = 1$$

Finding the probabilities for continuous outcomes:

- is represented as area under a curve,
- is in the context of a range of values, and
- the probability of hitting an exact value is 0

Example. Suppose a coffee shop has done extensive research and knows each customer is helped in under 5 minutes. The shaded area of the graph represents the probability that a customer will wait less than 2 minutes.

Example. Suppose a bus arrives at the bus stop every 12 minutes. If you arrive at the bus stop at a randomly chosen time, then the probability distribution for the number of minutes you must wait is shown in the graph below:



$h = \frac{1}{12}$

$w = 12 - 0 = 12$

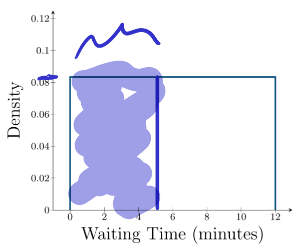
Area = 1

$A = wh$

$\frac{1}{12} = \frac{12h}{12}$

$\boxed{\frac{1}{12} = h}$

Find the probability that you will have to wait less than 5 minutes.



$\frac{1}{12}$

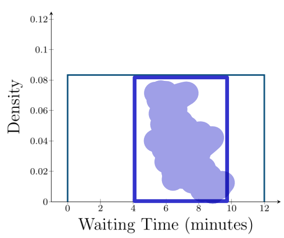
Wait time

$P(X < 5) = P(X \le 5) = (5-0)\frac{1}{12} = \frac{5}{12}$ $w \cdot h$

$= \boxed{0.41\overline{6}}$

Only equal with cont. prob.

Find the probability that you will have to wait between 4 and 10 minutes.

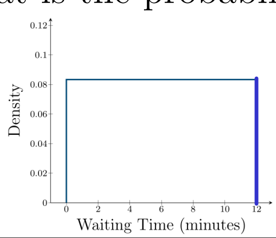


$P(4 < X < 10) = P(4 \le X \le 10)$

$= (10-4) \cdot \frac{1}{12}$ $w \cdot h$

$= \frac{6}{12} = \boxed{\frac{1}{2} = 0.5}$

What is the probability that you will have to wait *exactly* 12 minutes?



$P(X = 12) = (12-12) \cdot \frac{1}{12}$ $w \cdot h$

$= 0 \cdot \frac{1}{12}$

$= \boxed{0}$

$P(X = 5) = 0$

6.2: The Normal Model

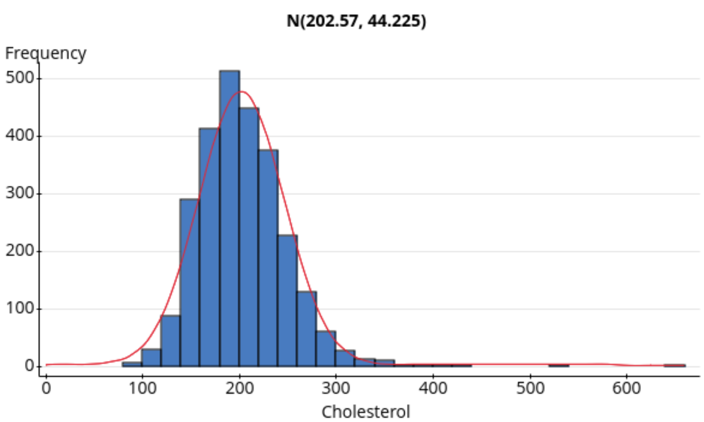
Definition.
The **Normal Distribution** is a symmetric, unimodal model that provides a very close fit for many numerical variables:

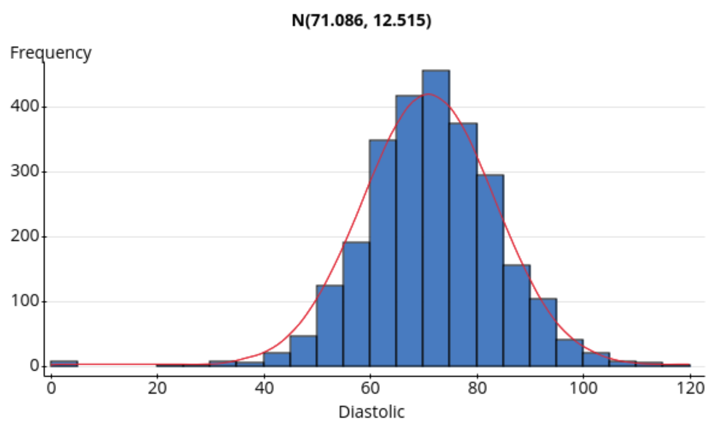
$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We use $N(\mu, \sigma)$ to denote the Normal Distribution with mean μ and standard deviation σ.
Note:

- μ and σ are used in the context of a probability distribution, whereas $\overline{x}$ and s are used for data.
- Other sources denote the Normal Distribution with mean μ and *variance* σ^2 as $N(\mu, \sigma^2)$ or $\mathcal{N}(\mu, \sigma^2)$.

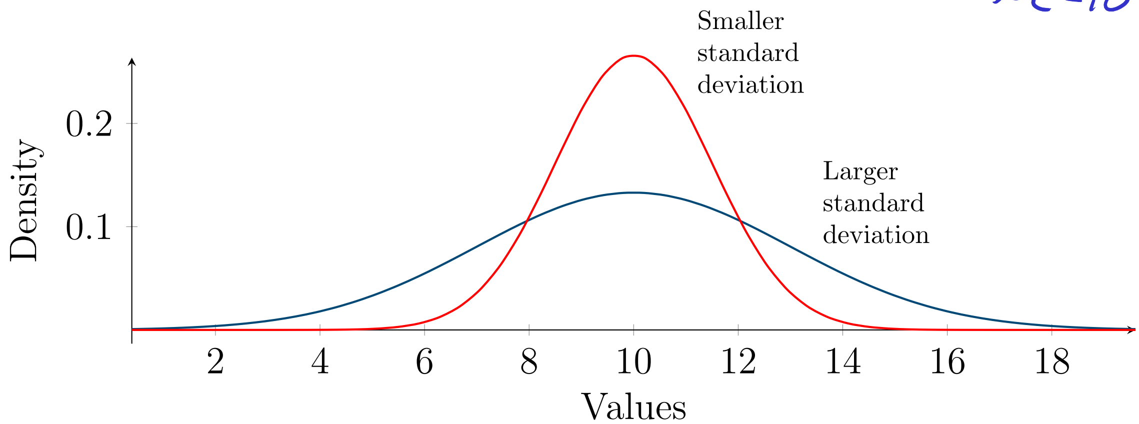
Example. Below are some histograms from a dataset that show the measured cholesterol and diastolic blood pressure from 2,793 people. These histograms have the Normal Distribution with the corresponding mean μ and standard deviation σ overlayed:



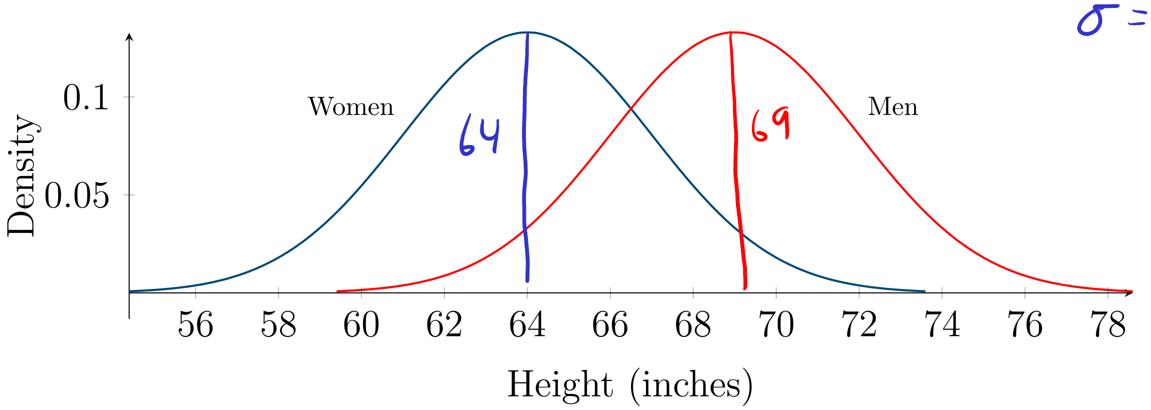


Example. Below is the graph of two Normal Distributions with equal means, but different standard deviations.

$\mu = 10$



Example. Below is the graph of two Normal Distributions with equal $\underbrace{\text{standard deviations}}_{\sigma = 3}$, but different means.

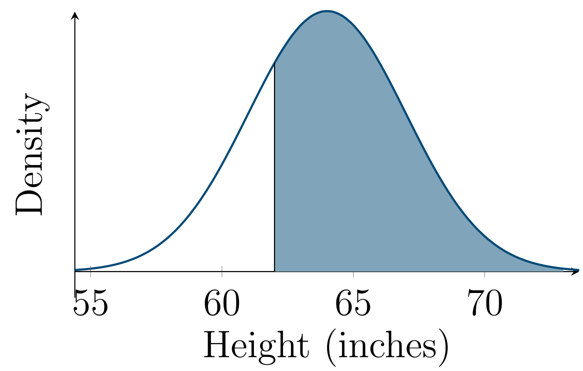


What is the area under each of the curves above?

1

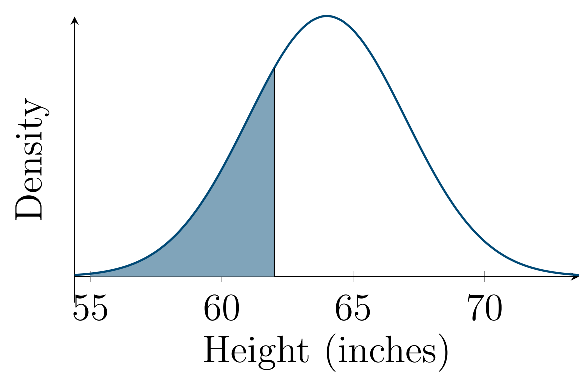
Example. Suppose that the Normal model $N(\overset{\mu}{64}, \overset{\sigma}{3})$ gives a good approximation to the distribution of adult women's height in the United States. If a women is chosen at random, what is the probability that

she is taller than 62"?



$P(\overset{\text{height}}{X} \geq 62) = 0.7475$

she is shorter than 62"?

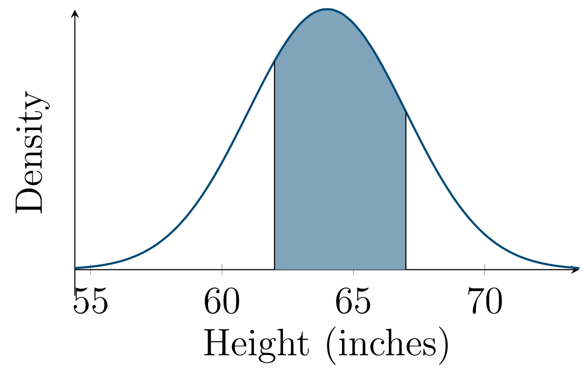


$P(X \leq 62) = 0.2525$

inc. order

$P(62 \le X \le 67) = \boxed{0.5889}$

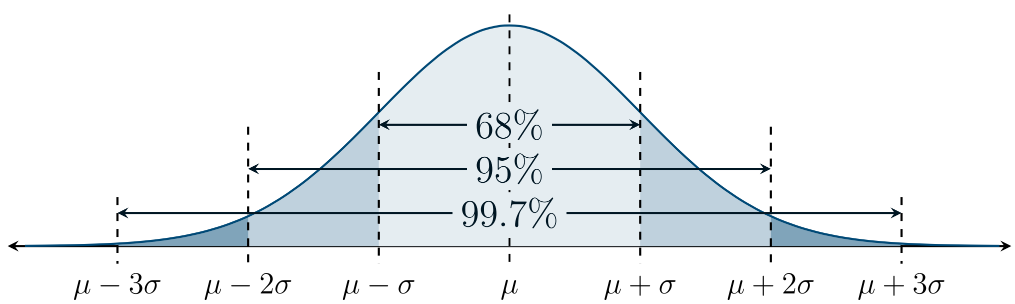
her height is between 62" and 67"?

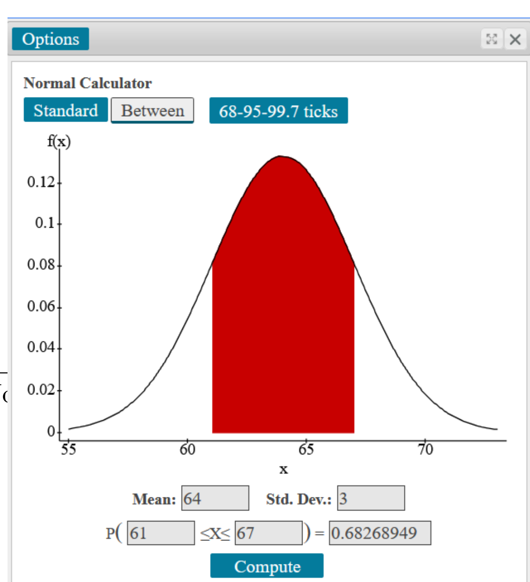


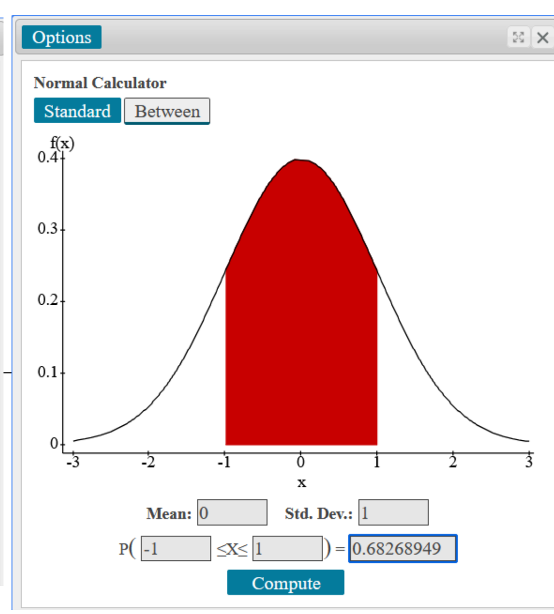
$P(62 \le X \le 67) = P(X \le 67) - P(X \le 62)$

$0.5889 = 0.8413 - 0.2525$

Example. "Verify" the empirical rule by using technology to find the probability that an observation lies within 1, 2, and 3 standard deviations.







Definition.
The **Standard Normal Distribution** is a $N(0,1)$:

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

We use the Standard Normal Distribution in conjunction with z-scores to compute probabilities:

$$z = \frac{X-\mu}{\sigma}$$

Example. Suppose the length of a newborn seal pup follows a Normal Distribution with a mean length of 29.5, and standard deviation 1.2. Solve the following by finding the z-score and then using a z-score table to compute the probability that a seal pup's length is

shorter than $\overbrace{28}^{X}$”,

$$z = \frac{28-29.5}{1.2} = -1.25$$

-1.2 0.05

$$P(X \le 28) = P(Z \le -1.25) = 0.10565$$

longer than 31”, and

$$z = \frac{31-29.5}{1.2} = 1.25$$

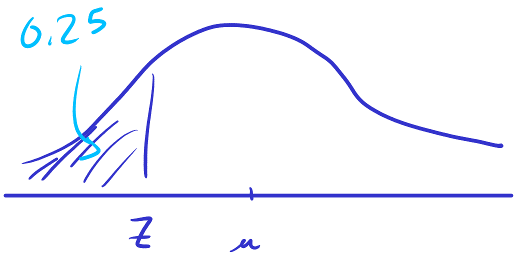
$$\begin{aligned} P(X \ge 31) &= P(Z \ge 1.25) \\ &= 1 - P(Z \le 1.25) \\ &= 1 - 0.89435 \\ &= 0.10565 \end{aligned}$$

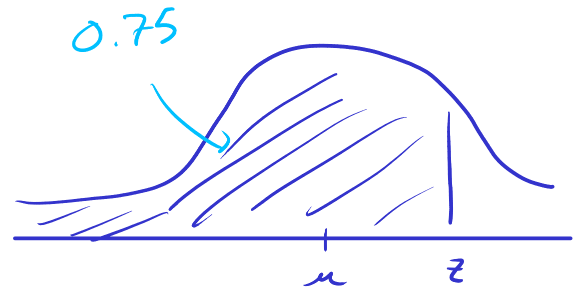
is between 28” and 31”.

$$\begin{aligned} P(28 \le X \le 31) &= P(-1.25 \le Z \le 1.25) \\ &= P(Z \le 1.25) - P(Z \le -1.25) \\ &= 0.89435 - 0.10565 \\ &= \boxed{0.7887} \end{aligned}$$

Example. Assume that women's heights follow a Normal distribution with mean 64" and standard deviation 3". Find the 25th and 75th percentile using

technology and





$P(X \le x) = 0.25$

$\Rightarrow \boxed{x = 61.9765}$

"Q_1"

$P(X \le x) = 0.75$

$\Rightarrow \boxed{x = 66.0235}$

"Q_3"

by hand.

$P(Z \le z) = 0.25$

$\Rightarrow z = -0.67$

$P(Z \le z) = 0.75$

$\Rightarrow z = 0.67$

$$\boxed{Z = \frac{X-\mu}{\sigma}}$$

$$(3)\,-0.67 = \frac{X-64}{3}(3)$$

$$+64 \quad -2.01 = X - 64 \; +64$$

$$\boxed{61.99 = X}$$

$$(3)\,0.67 = \frac{X-64}{3}(3)$$

$$+64+ \; 2.01 = X - 64 \; +64$$

$$\boxed{66.01 = X}$$